Understanding User Behavior in Online Social Networks: A Survey

Long Jin, University of California, San Diego
Yang Chen, Duke University
Tianyi Wang, Tsinghua University
Pan Hui, Hong Kong University of Science and Technology/Telekom Innovation Laboratories
Athanasios V. Vasilakos, Kuwait University

ABSTRACT

Currently, online social networks such as Facebook, Twitter, Google+, LinkedIn, and Foursquare have become extremely popular all over the world and play a significant role in people's daily lives. People access OSNs using both traditional desktop PCs and new emerging mobile devices. With more than one billion users worldwide, OSNs are a new venue of innovation with many challenging research problems. In this survey, we aim to give a comprehensive review of state-of-the-art research related to user behavior in OSNs from several perspectives. First, we discuss social connectivity and interaction among users. Also, we investigate traffic activity from a network perspective. Moreover, as mobile devices become a commodity, we pay attention to the characteristics of social behaviors in mobile environments. Last but not least, we review malicious behaviors of OSN users, and discuss several solutions to detect misbehaving users. Our survey serves the important roles of both providing a systematic exploration of existing research highlights and triggering various potentially significant research in these topics.

INTRODUCTION

In recent years, online social networks (OSNs) have dramatically expanded in popularity around the world. According to the data in October 2012, Facebook has 1.01 billion people using the site each month.[1] Moreover, the numbers of users in five popular OSNs are listed in Table 1. The rapid growth of OSNs has attracted a large number of researchers to explore and study this popular, ubiquitous, and large-scale service. In this article, we focus on understanding user behavior in OSNs.

OSN user behavior covers various social activities that users can do online, such as friendship creation, content publishing, profile browsing, messaging, and commenting. Notably, these activities can be legitimate or malicious. Understanding OSN user behavior is important to different Internet entities in several aspects:

- For Internet service providers (ISPs), as OSN traffic is growing quickly and becoming significant, they want to learn the evolution of the traffic pattern of OSNs. This can guide them to do some infrastructural actions (e.g., adding traffic optimization in network middle-boxes).
- For OSN service providers, it helps them understand their customers' attitudes toward different functions, especially for some experimental functions. Moreover, from the perspective of infrastructure investment, such as which locations are most cost-effective to build data centers or which content delivery network (CDN) cluster could be leveraged to deliver frequently accessed data, understanding users' geographic distribution and traffic activity is vital.
- For OSN users, behavior study is important to enhance user experience. For example, there are numerous malicious accounts in OSNs. These accounts generate unwanted messages for legitimate users. Therefore, identifying and blocking malicious users are very important to ensure good user experience.

Our survey contains four aspects of understanding user behavior in OSNs. First, a social graph is a classic and effective mathematical model to represent the relationship between users in OSNs, and has been widely used in OSN research. Based on four different types of social graphs, we discuss the aspect of *connectivity and interaction*. Second, network monitoring records detailed traffic activity of OSNs and provides us with a method to understand the network usage of OSNs. Also, network-based measurement results can demonstrate more users' activities than using the social graph only. Therefore, we focus on the perspective of *traffic activity*. Third, the rapid development of mobile platforms and applications plays an important role in OSN-related applications. Mobile devices not only

[1] *http://finance.yahoo.com/news/number-active-users-facebook-over-years-214600186—finance.html*

provide a venue for users to access OSNs everywhere, but also establish mobile-centric functions like location-based service (LBS) in OSNs. Understanding the behavior of mobile users can be leveraged to enhance the performance of mobile social applications and systems. Thus, we review the studies of *mobile social behavior*. Last but not least, OSNs introduce new challenges related to security and privacy. Malicious behaviors, such as spam and Sybil attacks, take place in OSNs and bring severe security threats. We show studies of *malicious behavior*.

OSN site	No. of users
Facebook	1.01 billion (Oct. 2012)
Twitter	500 million (Apr. 2012)
Google+	400 million (Sep. 2012)
LinkedIn	175 million (Jun. 2012)
Foursquare	25 million (Sep. 2012)

Table 1. *Information about five popular OSNs.*

CONNECTIVITY AND INTERACTION

MOTIVATION AND CHALLENGES

The social graph is an effective and widely-used mathematical tool to represent the relationships among users in OSNs, which benefits the analysis of social interactions and user behavior characterization. Usually, social networks can be modeled as undirected graphs (e.g., friendship graph, interaction graph) or directed graphs (e.g., latent graph, following graph) according to the properties of OSNs. Table 2 lists four different types of social graphs. Based on these graph types, we discuss the connectivity and interaction among OSN users. Moreover, the huge size of the social graph challenges the effectiveness of analysis. Thus, graph sampling and crawling techniques have been proposed to deal with this problem. In this section, we investigate several measurement, analysis, and modeling works related to the social graph.

Type	Edge
Friendship graph	Friendship between users
Interaction graph	Visible interaction, such as posting on a wall
Latent graph	Latent interaction, such as browsing profile
Following graph	Subscribe to receive all messages

Table 2. *Four different types of social graph.*

EXISTING SOLUTIONS AND DISCUSSION

Undirected Graph Model — For a *friendship graph*, every user is denoted as a node, and the friendship between any user pair is represented by an edge. Wilson *et al.* [1] try to find out whether social links are valid indicators of user interactions. They define wall posts and photo comments as interactions. Based on the crawled data from Facebook, they have found that users tend to interact mostly with only a small subset of their friends, while often having no interaction with up to half of their friends. Therefore, friendship in OSNs can hardly be viewed the same as friendship in the real world. Correspondingly, a new *interaction graph* is proposed to reflect the real user interactions in social networks, where only visible interaction between two users can create an edge in the graph, instead of being friends only. Using two representative applications, spam and Sybil protections, they demonstrate that using an interaction graph performs better than using a friendship graph.

Directed Graph Model — *Latent interactions* are passive actions of OSN users (e.g., profile browsing) that cannot be observed by traditional measurement techniques. Jiang *et al.* [2] study latent interactions based on the crawled data of Renren, the largest OSN provider in China. Renren tracks the most recent nine visitors to every user's profile, making the measurement of latent interactions possible. In a directed *latent graph*, a directed edge from A to B indicates A has visited B's profile. Therefore, the in-degree of a node shows the number of visitors to that user's profile, while the out-degree reveals the number of profiles that user has visited. A comparison between latent interactions and visible interactions is conducted based on Renren's crawled data, which contains 42 million users and 1.66 billion social links. There are three major findings. First, latent interactions are significantly more prevalent and frequent than visible interactions. Second, latent interactions are non-reciprocal in nature. Last but not least, the profile popularity is uncorrelated with the frequency of content updates or number of friends for very popular users. The characteristics of latent graphs are shown to fall between visible interaction graphs and classical friendship graphs.

Hwak *et al.* [3] perform extensive measurement on Twitter, the world's largest microblogging service, and reveals its power in information spreading on the news media level. In Twitter's *following graph*, a directed edge from A to B indicates A has subscribed to receive B's latest messages. The collected data is crawled over 24 days, with 41.7 million user profiles, 1.47 billion relations, 4262 trending topics, and 106 million tweets. It introduces a directed graph model to give a basic informative overview of Twitter, studies the distribution of followers/followees, and analyzes how the number of followers or followees affects the number of tweets. Additionally, in order to show how Twitter acts as a social medium and top users influence other users, this article tries to rank the users by number of followers, page rank, and retweets. The rankings by number of followers and page rank are almost the same, and the top users in the rankings are either celebrities or news media accounts. This article also analyzes the trending topics in Twitter and compares it with other media. It is found

A fast increase in the number of users makes the size of social graphs larger and larger, which presents researchers with a big challenge when performing any analysis with limited computation and storage capability. Graph sampling techniques are used to get a smaller but representative snapshot of social graphs.

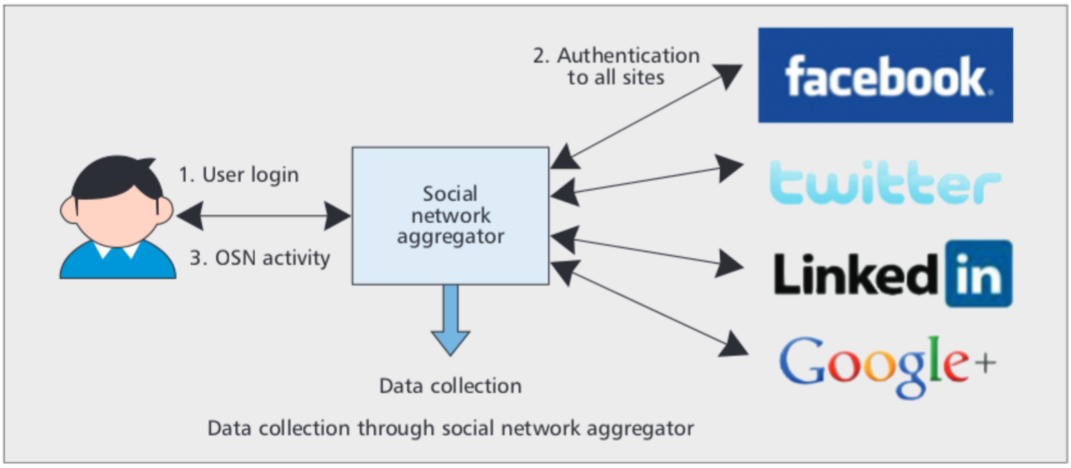

Figure 1. *Data collection through a social network aggregator.*

that the majority (over 85 percent) of trending topics in Twitter are headline or persistent news in nature, which reveals Twitter's live broadcasting nature and confirms Twitter's role as a news medium.

Graph Sampling — A fast increase in the number of users makes the size of social graphs larger and larger, which presents researchers with a big challenge when performing any analysis with limited computation and storage capability. Graph sampling techniques are used to get a smaller but representative snapshot of social graphs, which preserves properties such as degree distribution. As shown in [4], the sampling result of Breadth-First Sampling (BFS) and Random Walk (RW) are biased toward high-degree vertices, although they have been widely used in social graph analysis. The Metropolis-Hasting RW (MHRW) and a Re-Weighted RW (RWRW) are proposed and proved to perform uniformly in sampling Facebook. The article also introduces online convergence diagnostics to assess sample quality during the sampling process. Frontier Sampling (FS) [5], which leverages multidimensional RW, is proposed to achieve lower estimation errors than RW, especially in the presence of disconnected or loosely connected graphs. Ribeiro *et al.* [5] show that FS is more suitable for estimating the tail of degree distribution than random vertex sampling. Moreover, FS can be made fully distributed without any coordination costs.

Future Work

The dynamic feature is an important aspect to deeply understand an OSN's user behavior. Much of the existing work tries to investigate an OSN in a relatively static way, by collecting or studying a static snapshot dataset. However, the growth of OSNs is extremely rapid. Every day new users join OSNs, while existing users make new friends or end social connections, join or leave groups, and so on. Considering this dynamic can extract more inherent information than studying static data, not only revealing the situation at a certain time but also predicting some future activities. Also, studying different time intervals and time granularities would lead to more interesting findings. There are several challenges for performing dynamic analysis. One is fast data collection and timely processing, where an unbiased and efficient graph sampling algorithm can play an important role. Also, collecting dynamic data raises challenges for information storage; therefore, the temporal and spatial dependence between different data items can be utilized for better compression.

Traffic Activity

Motivation and Challenges

Different kinds of social graphs can reveal how users connect and interact with each other. However, due to the limited information that the graph can represent, various types of users' activities cannot be characterized (e.g., time duration of browsing a profile). An observation from network operators can monitor such information easily, and interpret how users use OSNs better. Furthermore, for ISPs, they have strong incentive to get better understanding of how the traffic pattern between end users and OSN sites will evolve, and take optimization actions according to the distribution and activities of OSN users. In this section, we review OSN user behavior study from the perspective of network traffic analysis.

Existing Solutions and Discussion

Traffic Monitoring — Besides crawling, people can also study OSNs by monitoring the corresponding network traffic. Benevenuto *et al.* [6] analyze the user behavior of OSNs based on detailed clickstream data obtained from a social network aggregator, as illustrated in Fig. 1.

In [6], the clickstream data was collected over 12 days with HTTP sessions of 37,024 users who accessed popular social networks. This article defines and analyzes the OSN session characteristics:
- The frequency of accessing OSNs
- Total time spent on OSNs
- Session duration of OSNs

Through the clickstream data, user activities are also identified. Forty-one types of user activities are classified into nine groups, and the popularity of different activities and the traffic bytes are analyzed. Interestingly, it is found that silent or latent interactions such as browsing account for more than 90 percent of user activities. Also, they show how users have different activities in different OSNs. They also characterize how users transit from one activity to another using a first-order Markov chain.

Schneider *et al.* [7] also study clickstream data, but their focuses are feature popularity, session characteristics, and the dynamics within OSN sessions. The distribution of HTTP request-response pairs reveal the popularity of different features. The popularity of features can be different among users from different areas and of different OSNs. It can also differ by the time spent by the users. Besides, the distribution of transmission bytes per OSN session is given, which helps the ISPs learn the traffic pattern of different OSNs. Photo features account for most traffic bytes of OSNs. It also shows the duration of sessions and number of subsessions within a session. Moreover, the article reveals the dynamics within OSN sessions. It is found that most users access web sites other than OSNs during OSN sessions for more than 1 min. That is, users can be inactive when accessing the OSNs.

Clickstream data contributes a lot in the user behavior study of OSNs. However, it can be incomplete, which restricts its usage and performance. First, click-stream data is limited by the collection duration, and the behavior of inactive users in the duration is not monitored. Moreover, the data is restricted by the monitoring locations. That is, only the behavior of users using certain monitored ISPs is captured.

Locality of Interest — Facebook is heavily dependent on centralized U.S. data centers to provide consistent service to users all over the world. Therefore, users outside the United States experience slow response time. Also, a lot of unnecessary traffic is generated on the Internet backbone. Wittie *et al.* [8] investigate the detailed causes of these two problems and identify mitigation opportunities. It is found that OSN state is amenable to partitioning, and its fine-grained distribution and processing can significantly improve performance without loss in service consistency. Based on simulations of reconstructed Facebook traffic over measured Internet paths, it is shown that user requests can be processed 79 percent faster and use 91 percent less bandwidth. Therefore, the partitioning of OSN state is an attractive scaling strategy for OSN service providers.

Navigation Characteristics — Nowadays, OSNs represent a significant portion of web traffic, comparable with search engines. Dunn *et al.* [9] try to understand the similarities and differences in the web sites users visit through OSNs vs. through search engines. Using web traffic logs from 17,000 digital subscriber line (DSL) subscribers of a Tier 1 ISP in the United States, it is found that OSN visitors are less likely to navigate to external web sites. But when they visit external web sites, OSN users will spend more time at those web sites compared to search engine users. Also, OSNs direct visitors to a narrower subset of the web than search engines. While web sites related to games and video are more commonly visited from OSNs, shopping and reference sites are common for search engines. Finally, OSNs send users to less popular domains more often than search engines. These findings can be useful to ISPs in network provisioning and traffic engineering.

Future Work

Most existing measurement and analysis projects are led by either academic groups or ISPs, without the active involvement of OSN service providers. Such a situation limits the insight of the study. On one hand, academic researchers always use extensive crawling to obtain the data, which encounters many restrictions from the OSN providers, such as traffic control (how many messages per IP and/or per account can be fetched in one hour). Also, some users may use privacy options to make their data unavailable. Last but not least, the huge number of users makes it almost impossible to get a timely snapshot, so data consistency cannot be guaranteed. On the other hand, although an ISP is able to capture and analyze all its traffic to/from an OSN site through traffic monitoring, it can only get a partial view of the whole site; that is, only users who get access to OSNs through a specific ISP's infrastructure can be observed. As we have discussed, user behavior study can be beneficial for OSN providers themselves. We envision that OSN providers can collaborate with academia and industrial researchers in order to understand user behavior in an insightful way. This can enhance the user experience interactively and quickly. Also, this will save operational costs for OSN providers.

We envision that OSN providers can collaborate with academia and industrial researchers in order to understand user behavior in an insightful way. This can enhance the user experience interactively and quickly. Also, this will save operational costs for OSN providers.

Mobile Social Behavior

Motivation and Challenge

Nowadays, due to the wide use of mobile devices, more and more web applications have been expanded to mobile platforms, as have OSN services. We believe that it is the right time to highlight the importance of mobile social networks (MSNs). In MSNs, mobile users can publish and share information based on the social connections among them. On one hand, most major OSN platforms such as Facebook, Twitter, and LinkedIn release mobile applications to allow users to access their services through mobile devices. On the other hand, more mobile-centric functions have been integrated into OSNs, such as location-based services and mobile communication. Understanding the user behavior in MSNs is very helpful for the design and implementation of MSN systems, improving the system efficiency in mobile environments or supporting better mobile-centric functions. In this section, we focus on studies of user behaviors in MSNs.

Existing Solutions and Discussion

Mobile Social Application — A large number of interesting and useful mobile social applications have been proposed. *Social Serendipity* [10] is a mobile-phone-based system that combines widely used mobile phones with the functionality

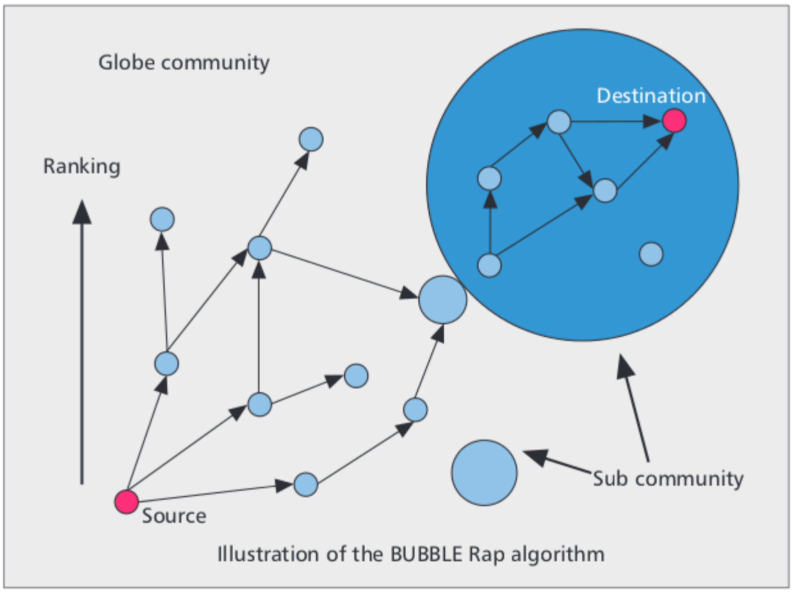

Figure 2. *Illustration of the BUBBLE Rap algorithm.*

of online introduction systems to cue informal face-to-face interactions between nearby users who do not know each other but probably should. Serendipity uses Bluetooth to sense nearby people and utilizes a centralized server to decide whether two users should be introduced to each other. The system calculates a similarity score by extracting the commonalities between two proximate users' profiles and behavioral data, and sums them according to user-defined weights. If the score is higher than the threshold set by both users, the system will inform them that someone nearby might be interested in them. For instance, internal collaboration in large companies can be facilitated by Serendipity for introducing people who are working on similar projects. It is emphasized that privacy issues are important and fundamental in Serendipity, and privacy-protecting tools should be designed carefully.

Geographical Prediction in OSN — Geography and social relationships are inextricably intertwined. As people spend more time online, data regarding these two dimensions are becoming increasingly precise, allowing building reliable models to describe their interaction. In [11], the study of user-contributed address and association data from Facebook shows that the addition of social information produces improvement in accuracy of predicting physical location. First, friendship as a function of distance and rank is analyzed. It is found that at medium to long-range distances, the probability of friendship is roughly proportional to the inverse of distance. However, at shorter ranges, distance does not influence much. Then the maximum likelihood approach is presented to predict the physical location of a user, given the known location of her friends. This method predicts the physical location of 69.1 percent of the users with 16 or more located friends to within 25 mi, compared to only 57.2 percent using IP-based methods.

Friendship and Mobility in LBSN —Although human movement and mobility patterns have a high degree of freedom and variation, they also exhibit structural patterns due to geographic and social constraints. Using cell phone location data, as well as data from two online location-based social networks, Cho *et al.* [12] aim to understand the basic laws that govern human motion and dynamics. It is found that humans experience a combination of strong short-range spatially and temporally periodic movement that is not impacted by the social network structure, while long-distance travel is more influenced by the social network ties. Furthermore, it is shown that social relationships can explain about 10 to 30 percent of all human movement, while periodic behavior explains 50 to 70 percent. Based on these findings, a model of human mobility is proposed that combines periodic short-range movements with travel due to the social network structure and gives an order of magnitude better performance than previous models.

Social-Based Routing in PSNs — Widely used smart devices with networking capability form novel networks, such as pocket switched network (PSN). Due to the mobility of devices, PSNs are intermittently connected, and effective routing protocols are essential in such networks. Previous methods relied on building and updating routing tables to deal with dynamic conditions. Actually, the social structure and the interaction of users of smart devices have a great influence on the performance of routing protocols. *BUBBLE Rap* [13] is a social-based forwarding method for PSNs. Two social and structural metrics, centrality and community, are used to effectively enhance delivery performance. As shown in Fig. 2, BUBBLE Rap first uses a centrality metric to spread out the messages (i.e., sending messages to more popular nodes), and then uses a community metric to identify the destination community and focus the messages to the destination. The evaluation shows that BUBBLE Rap has a similar delivery ratio, but much lower resource utilization than flooding, control flooding, and other social-based forwarding schemes.

Content Distribution in MSN — Ioannidis *et al.* [14] study the dissemination of dynamic content, such as news and traffic information, over an MSN. In this application, mobile users subscribe to a dynamic-content distribution service offered by their service provider. To improve coverage and increase capacity, it is assumed that users share any content updates they receive with other users they meet. Reference 14 determines how the service provider can allocate its bandwidth optimally to make the content at users as "fresh" as possible. Moreover, there is a condition under which the system with high scalability is specified: even if the total bandwidth dedicated by the service provider remains fixed, the expected content age at each user grows slowly (as $\log(n)$) with the number of users n.

FUTURE WORK

There are several fundamental issues that require continuous exploration in the research related to user behavior in MSNs, including incentive mechanism, identity management, trust, reputation and privacy, energy efficiency, methods for social network metrics estimation and community detection, content distribution and sharing protocols, and precise localization techniques for geographic and semantic spaces. A comprehensive summary related to applications, architectures, and protocol design issues for MSNs can be found in [15].

Furthermore, we believe social data delivery and social applications in mobile environments rouse challenges in several layers of the Internet protocol stack. Let us list three examples here. First, we need a better transport layer protocol to handle packet loss caused by wireless environment and host mobility. Second, to efficiently deliver popular content desired by multiple MSN users, we need to deploy social-aware proxies in the network infrastructure to eliminate duplicate transmission. The deployment of those proxies needs to carefully consider social connections, users' geolocations, and the topology of the underlying wired/wireless Internet. Third, context-aware services will become very useful in MSNs. Such services will let user express their demand for social activities in cyberspace in a human-readable fashion, thus making social interaction among mobile users easier. All three of these examples need lots of work in data analyzing, modeling, and prototyping.

MALICIOUS BEHAVIOR

MOTIVATION AND CHALLENGES

The usage of OSNs introduces numerous security and privacy threats. For instance, as a user needs to interact with other users through an OSN service provider, its activities and uploaded data can be tracked and stored by the OSN service provider. These data (photos, articles, public posts, private messages, etc.) may be leaked to a third party without the user's explicit authorization, even when the user regards some of these as confidential. Moreover, Sybil attacks are very common in OSNs, as a user can register multiple fake accounts maliciously. These fake accounts can perform various malicious activities including spamming, obtaining privacy contact lists, misleading crowd-sourcing results, and so on. Besides those, Gao *et al.* [16] list several other attacks such as re-identification and de-anonymization of anonymized OSN data, fetching personal data through untrusted third-party applications, cross-site profile cloning, social spamming, and phishing. Due to space limitation, this survey mainly focuses on malicious behavior in OSNs, including spam and Sybil attacks.

EXISTING SOLUTIONS AND DISCUSSION

Social Spam — OSNs are popular collaboration tools for millions of users and their friends. Unfortunately, they also become effective tools for executing spam campaigns and spreading malware. Intuitively, a user is more likely to respond to a message from a friend than from a stranger; thus, social spamming is a more effective distribution mechanism than traditional email. Gao *et al.* [17] study a large dataset composed of over 187 million wall messages among 3.5 million Facebook users. The system detected 200,000 malicious wall posts with embedded URLs originating from more than 57,000 accounts. It is shown that more than 70 percent of all malicious wall posts advertise phishing sites. It is also found that more than 97 percent are compromised accounts rather than"fake" accounts created solely for the purpose of spamming. Finally, spamming dominates actual wall post activity in the early morning hours, when normal users are asleep.

Lumezanu *et al.* [18] perform a joint analysis of spam in email and social networks. Spam data from Yahoo's web-based email service and Twitter are used to characterize the publishing behavior and effectiveness of spam advertised across both platforms. It is shown that email spammers that also advertise on Twitter tend to send more email spam than those advertising exclusively through email. Furthermore, sending spam on both email and Twitter has better exposure than spamming exclusively with email: spam domains appearing on both platforms are looked up by an order of magnitude more networks than domains using just one platform.

Social-Graph-Based Sybil Defense — Sybil attacks are the fundamental problem in peer-to-peer and other distributed systems. In a Sybil attack, a malicious attacker creates multiple fake identities to influence the working of systems that depend on open membership, such as recommendation and delivery systems. Recently, a number of social network-based schemes, such as *SybilGuard*, *Sybillimit*, *SybilInfer*, and *SumUp*, have been proposed to mitigate Sybil attacks. Viswanath *et al.* [19] develop a deep understanding of these approaches. It shows that existing Sybil defense schemes, which can be viewed as graph partitioning algorithms, work by identifying local communities (i.e., clusters of nodes more tightly knit than the rest of the graph) around a trusted node. Therefore, the substantial amount of prior research on general community detection algorithms can be used to design effective and novel Sybil defense schemes.

Usually, binary Sybil/non-Sybil classifiers have high false positives; thus, manual inspection needs to be involved in the decision process for suspending an account. *SybilRank* [20] aims to efficiently derive a Sybil-likelihood ranking; only the most suspicious accounts need to be inspected manually. It is based on efficiently computable early-terminated RWs and is suitable for parallel implementation on a framework such as Map Reduce, uncovering Sybils in OSNs with millions of accounts. SybilRank is deployed and tested in the operation center of Tuenti, which is the largest OSN in Spain with 11 million users. Almost 100 and 90 percent of the 50K and 200K accounts, which SybilRank regards as the most suspicious, are indeed fake. In contrast, the hit rate of the current user-report-based approach is only 5 percent. Thus, SybilRank represents a significant step toward practical Sybil defense.

FUTURE WORK

Because of the abundance of available personal information, OSNs suffer a vital problem of privacy breach, and such attacks may be caused by

Because of the abundance of available personal information, OSNs suffer from the vital problem of privacy breach. Such attacks may be caused by three primary parties in the OSN: service providers, malicious users, and third-party applications.

We envision that this research line will enhance the user experience from various aspects, as well as satisfy different players, including the infrastructure providers, service providers, and end users.

three primary parties in the OSN: service providers, malicious users, and third-party applications. Decentralized OSN is a potential architecture to protect sensitive information from leaking out to service providers and third-party applications. However, how to provide incentives to encourage users to switch to a decentralized OSN is challenging, especially for users who do not care much about their privacy and security. Furthermore, the recipients of shared information should be controlled by the users themselves. Instead of sharing information based on the virtual links in OSNs, real-life relationship between users should also be taken into account. Finally, Sybil defense is still a hot topic and more solid work are expected to conduct in this area. We foresee that semantic information extracting from user profiles and social behavior can be used for Sybil detection, and should be utilized collaboratively with existing schemes.

Conclusion

In this survey, we study user behavior in OSNs from four different perspectives: connection and interaction, traffic activity, mobile social behavior, and malicious behavior. We review the existing representative schemes and also provide potential future directions. We envision that this research line will enhance the user experience from various aspects, as well as satisfy different players, including the infrastructure providers, service providers, and end users. We believe that further research of user behavior in OSNs will generate more interesting research problems and exciting solutions in this area.

References

[1] C. Wilson *et al.*, "User Interactions in Social Networks and Their Implications," *Proc. EuroSys*, 2009.

[2] J. Jiang *et al.*, "Understanding Latent Interactions in Online Social Networks," *Proc. IMC*, 2010.

[3] H. Kwak *et al.*, "What Is Twitter, a Social Network or a News Media?," *Proc. WWW*, 2010.

[4] M. Gjoka *et al.*, "Practical Recommendations on Crawling Online Social Networks," *IEEE Trans. Commun.* Special Issue on Measurement of Internet Topologies, vol. 29, no. 9, Oct. 2011.

[5] B. Ribeiro and D. Towsley, "Estimating and Sampling Graphs with Multidimensional Random Walks," *Proc. IMC*, 2010.

[6] F. Benevenuto *et al.*, "Characterizing User Behavior in Online Social Networks," *Proc. IMC*, 2009.

[7] F. Schneider *et al.*, "Understanding Online Social Network Usage from a Network Perspective," *Proc. IMC*, 2009.

[8] M. Wittie *et al.*, "Exploiting Locality of Interest in Online Social Networks," *Proc. CoNext*, 2010.

[9] C. W. Dunn *et al.*, "Navigation Characteristics of Online Social Networks and Search Engines Users," *Proc. WOSN*, 2012.

[10] N. Eagle and A. Pentland, "Social Serendipity: Mobilizing Social Software," *IEEE Pervasive Computing*, vol. 4, no. 2, 2005, pp. 28–34.

[11] L. Backstrom, E. Sun, and C. Marlow, "Find Me If You Can: Improving Geographical Prediction with Social and Spatial Proximity," *Proc. WWW*, 2010.

[12] E. Cho, S. Myers, and J. Leskovec, "Friendship and Mobility: User Movement in Location-Based Social Networks," *Proc. KDD*, 2011.

[13] P. Hui, J. Crowcroft, and E. Yoneki, "BUBBLE Rap: Social-Based Forwarding in Delay Tolerant Networks," *IEEE Trans. Mobile Computing*, vol. 10, Nov. 2011, pp. 1576–89.

[14] S. Ioannidis, A. Chaintreau, and L. Massoulie, "Optimal and Scalable Distribution of Content Updates over a Mobile Social Network," *Proc. INFOCOM*, 2009.

[15] N. Kayastha *et al.*, "Applications, Architectures, and Protocol Design Issues for Mobile Social Networks: A Survey," *Proc. IEEE*, vol. 99, no. 12, 2011, pp. 2130–58.

[16] H. Gao *et al.*, "Security Issues in Online Social Networks," *IEEE Internet Computing*, vol. 15, no. 4, 2011.

[17] H. Gao *et al.*, "Detecting and Characterizing Social Spam Campaigns," *Proc. IMC*, 2010.

[18] C. Lumezanu and N. Lumezanu, "Observing Common Spam in Tweets and Email," *Proc. IMC*, 2012.

[19] B. Viswanath *et al.*, "An Analysis of Social Network-Based Sybil Defenses," *Proc. SIGCOMM*, 2010.

[20] Q. Cao *et al.*, "Aiding the Detection of Fake Accounts in Large-Scale Online Social Services," *Proc. NSDI*, 2012.

Biographies

LONG JIN longjin@cs.ucsd.edu) is currently a Ph.D. student in the Department of Computer Science and Engineering, University of California, San Diego. He received his B.S. and M.S. degrees in Electronic Engineering from Tsinghua University in 2010 and 2013, respectively. He visited Carnegie Mellon University and Microsoft Research Asia in 2012. His research interests include social networks, mobile computing and wireless networks.

YANG CHEN (ychen@cs.duke.edu) is a postdoctoral associate in the Department of Computer Science, Duke University. From September 2009 to March 2011, he was a research associate at the University of Goettingen, Germany. He received his B.S. and Ph.D. degrees from the Department of Electronic Engineering, Tsinghua University, in 2004 and 2009, respectively. He visited Stanford University (in 2007) and Microsoft Research Asia (2006-2008) as a visiting student. His research interests include Internet architecture and protocols, cloud computing, and online/mobile social networks.

TIANYI WANG (wty07@mails.tsinghua.edu.cn) is now pursuing his Ph.D. degree in the Department of Electronic Engineering, Tsinghua University, and his advisor is Prof. Xing Li. He received his B.S. degree in electronic engineering from Tsinghua University in 2011. His research interests include analysis of online social networks and data mining.

PAN HUI (panhui@cse.ust.hk) received his Ph.D degree from the Computer Laboratory, University of Cambridge, and earned his M.Phil. and B.Eng. from the Department of Electrical and Electronic Engineering, University of Hong Kong. He is currently a faculty member of the Department of Computer Science and Engineering at the Hong Kong University of Science and Technology where he directs the System and Media Lab. He also serves as a Distinguished Scientist of Telekom Innovation Laboratories (T-labs) Germany and an adjunct professor of social computing and networking at Aalto University, Finland. Before returning to Hong Kong, he spent several years at T-labs and Intel Research Cambridge. He has published more than 100 research papers, and has several granted and pending European patents. He has founded and chaired several IEEE/ACM conferences/workshops, and served on the technical program committees of numerous international conferences and workshops including IEEE INFOCOM, SECON, MASS, GLOBECOM, WCNC, and ITC.

ATHANASIOS V. VASILAKOS (vasilako@sci.kuniv.edu.kw, vasilako@ath.forthnet.gr) is currently professor at Kuwait University. He has served or is serving as an Editor for many technical journals, such as *IEEE TNSM*, *IEEE TC*,*IEEE TSMC-PART B*, *IEEE TITB*, *ACM TAAS*, and *IEEE JSAC* Special Issues in May 2009, and January and March 2011. He is Chairman of the Council of Computing of the European Alliances for Innovation.